**PAPER TO BE PRESENTED AT ANNUAL MEETING OF AMERICAN NAME SOCIETY, NEW YORK 5 JAN 2019**

# What's in a Name? Issues in Named Entity Recognition

By: Brian D. Joseph D., Alexander H. Erdmann, Christopher G. Brown, Petra Ajaka, Benjamin O. Allen, Marie-Catherine H de Marneffe, Micha Elsner, Andrew B. Kessler, Colleen Kron, William L. Little, James C. Wolfe, Matias D. Grioni, Hannah F. Young (*The Herodotos Project*, Ohio State University)

With support from the National Endowment for the Humanities, an Ohio State University team is cataloguing the names of peoples (groups, tribes) mentioned in classical sources as part of the Herodotos Project, an ethnohistory project aimed at compiling all known information about these ancient peoples. To automatically identify names of these peoples and places, we are developing Named-Entity Recognition (NER) systems for Latin and Greek. NER is a machine learning technique that requires training on large amounts of text in which humans have manually, accurately, and consistently identified the named entities of interest.

We thus have been annotating different genres of Latin and Greek texts — poetry, literary prose, epistles, historiography — to develop a training corpus of identified names for the computational algorithm. This task of distinguishing personal names from group and place names has raised issues of wider interest for onomastic research:

i. Ultimately, every noun is construable as a name.
ii. Proper names are defined by a specificity that can be hard to identify.
iii. Many group names derive from place names, others from personal names.
iv. Personal names derive from places that are evocative rather than genealogical.
v. Other personal names derive from groups, e.g. names of Roman gentes.
vi. Place names themselves can derive from groups and even individuals.

Thus, annotators must decide whether specific names should be construed as groups, persons, or places according to the logic of our program's algorithms, or other historical criteria. We discuss these theoretical and practical problems and outline our solutions.